



UNIVERSITAT OBERTA DE CATALUNYA (UOC)

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (*Data Science*)

TRABAJO FINAL DE MÁSTER

ÁREA: 3

Generación Asistida por IA de Terrenos 3D y Ecosistemas Realistas

Un Enfoque Basado en Aprendizaje Profundo para la Creación de Paisajes Sintéticos

Autor: Ouassim Aouattah Akandouch

Tutor: Diego Calvo Barreno

Profesor: Laia Subirats Maté

Barcelona, 19 de diciembre de 2023

Créditos/Copyright

Una página con la especificación de créditos/copyright para el proyecto (ya sea aplicación por un lado y documentación por el otro, o unificadamente), así como la del uso de marcas, productos o servicios de terceros (incluidos códigos fuente). Si una persona diferente al autor colaboró en el proyecto, tiene que quedar explicitada su identidad y qué hizo.

A continuación se ejemplifica el caso más habitual, aunque se puede modificar por cualquier otra alternativa:



Esta obra está sujeta a una licencia de Reconocimiento - NoComercial - SinObraDerivada 3.0 España de CreativeCommons.

FICHA DEL TRABAJO FINAL

| | |
|-----------------------------------|--|
| Título del trabajo: | Generación Asistida por IA de Terrenos 3D y Ecosistemas Reales |
| Nombre del autor: | Ouassim Aouattah Akandouch |
| Nombre del colaborador/a docente: | Diego Calvo Barreno |
| Nombre del PRA: | Laia Subirats Maté |
| Fecha de entrega (mm/aaaa): | MM/AAAA |
| Titulación o programa: | Ciencia de Datos |
| Área del Trabajo Final: | TFM - Área 3 - Aula 1 |
| Idioma del trabajo: | Español |
| Palabras clave | Aprendizaje Profundo, Contenido Generativo, Entornos 3D |

Dedicatoria/Cita

Breves palabras de dedicatoria y/o una cita.

Agradecimientos

Si se considera oportuno, mencionar a las personas, empresas o instituciones que hayan contribuido en la realización de este proyecto.

Abstract

This research aims to develop an innovative AI-driven approach for generating 3D landscapes and realistic ecosystems. By utilizing deep learning models, the goal is to create a system capable of transforming textual descriptions into heightmaps and color maps, which describe the ecological features, for subsequent conversion into three-dimensional models using a 3D engine and procedural generation.

The project will focus on training an AI model with a dataset of real-world heightmaps and ecological data. This model will generate realistic terrains based on predefined parameters and understand and execute textual descriptions, making it accessible to users without technical expertise in 3D or AI.

This study will benefit various sectors such as video games, entertainment, urban planning, environmental conservation, and scientific research by offering a versatile tool for creating and analyzing realistic terrains and ecosystems.

Keywords: 3D Terrain Generation, GANs, Diffusion Models, Heightmaps, Ecological Data, Synthetic Data, Deep Learning, 3D Visualization, Memory, Master's Thesis.

Resumen

Esta investigación busca desarrollar un enfoque innovador mediante IA para la generación de paisajes 3D y ecosistemas realistas. Utilizando modelos e aprendizaje profundo, se pretende crear un sistema capaz de transformar descripciones textuales en mapas de alturas y de colores, que describen las características ecológicas de estos, para su posterior transformación en modelos tridimensionales mediante un motor 3D y generación procedimental.

El proyecto se centrará en entrenar un modelo de IA con un dataset de mapas de altura del mundo real y datos ecológicos. Este modelo generará terrenos realistas según parámetros predefinidos y comprenderá y ejecutará descripciones textuales, siendo accesible para usuarios sin conocimientos técnicos en 3D o IA.

Este estudio beneficiará a diversos sectores, como videojuegos, entretenimiento, planificación urbana, conservación ambiental e investigación científica, ofreciendo una herramienta versátil para la creación y análisis de terrenos y ecosistemas realistas.

Palabras clave: Generación de Terrenos 3D, GANs, Modelos de Difusión, Mapas de Altura, Datos Ecológicos, Datos Sintéticos, Aprendizaje Profundo, Visualización 3D, M, Memoria, Trabajo de Final de Grado/Máster.

Índice general

| | |
|--|-------------|
| Abstract | IX |
| Resumen | X |
| Índice | XI |
| Llistado de Figuras | XIII |
| Listado de Tablas | 1 |
| 1. Introducción | 2 |
| 1.1. Descripción de la propuesta | 2 |
| 1.1.1. Justificación | 2 |
| 1.1.2. Motivación Personal | 3 |
| 1.1.3. Objetivos | 3 |
| 1.1.4. Metodología | 4 |
| 1.1.5. Planificación | 4 |
| 2. Estado del Arte | 5 |
| 2.1. Introducción | 5 |
| 3. Fundamentos Teóricos | 10 |
| 3.1. Redes Generativas Antagónicas (GANs) | 10 |
| 3.1.1. Arquitectura y Funcionamiento | 10 |
| 3.1.2. Red Generativa Antagónica Convolucional Profunda (DCGANs) | 11 |
| 3.2. Pix2Pix: Entendiendo las Redes Condicional GANs | 12 |
| 3.2.1. Arquitectura del Modelo Pix2Pix | 12 |
| 3.2.2. Aplicaciones del Pix2Pix GAN | 12 |
| 3.2.3. Entrenamiento y Preprocesamiento | 12 |

| | |
|--|-----------|
| 4. Metodología y Desarrollo | 14 |
| 4.1. Implementación de la DCGAN para Mapas de Altura | 14 |
| 4.1.1. Preprocesado de Datos | 14 |
| 4.1.2. Inicialización de Pesos | 14 |
| 4.1.3. Generador | 14 |
| 4.1.4. Discriminador | 16 |
| 4.1.5. Funciones de pérdida y optimizadores | 16 |
| 4.1.6. Entrenamiento | 16 |
| 4.1.7. Resultados y Evaluación | 16 |
| 4.2. Implementación de Pix2Pix para Texturización | 16 |
| 4.2.1. Preprocesado de Datos | 17 |
| 4.2.2. Generador | 17 |
| 4.2.3. Discriminador | 18 |
| 4.2.4. Entrenamiento | 18 |
| 4.2.5. Resultados y Evaluación | 18 |
| Bibliografía | 18 |

Índice de figuras

| | | |
|------|---|----|
| 1.1. | Planificación Gantt | 4 |
| 2.1. | Estructura de Terrain-CGAN. | 7 |
| 2.2. | Proceso de generación de terrenos. | 8 |
| 2.3. | 3D-GPT emplea LLM como un sistema multiagente con tres agentes colaborativos para la generación de procedimientos 3D. | 9 |
| 4.1. | Datos de Entrenamiento | 15 |
| 4.2. | Estructura de un Generador | 15 |
| 4.3. | Métricas | 16 |
| 4.4. | Resultados | 17 |
| 4.5. | Ejemplo de una arquitectura U-Net básica (Ref:- https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/u-net-architecture.png) | 18 |
| 4.6. | Input, Label, Output | 19 |

Índice de cuadros

Capítulo 1

Introducción

1.1. Descripción de la propuesta

La propuesta se centra en desarrollar un sistema avanzado de Inteligencia Artificial (IA) capaz de generar terrenos 3D y ecosistemas realistas a partir de descripciones textuales y parámetros definidos por el usuario. Utilizaremos Redes Generativas Adversas (GANs) o tecnologías similares para modelar y generar datos sintéticos precisos. El sistema se entrenará utilizando un conjunto de datos comprensivo que incluirá mapas de altura del mundo real y datos asociados de la vegetación y la flora. A partir de los mapas de altura sintéticos se generará terrenos tridimensionales que podrán ser llenados de vegetación con métodos procedimentales a partir de los mapas de colores (datos ecológicos).

1.1.1. Justificación

1. **Crecimiento del interés:** Las herramientas de creación basadas en IA están inaugurando una nueva era de creatividad en los estudios que están dispuestos a aprovechar estas tecnologías. Se espera que la demanda de este tipo de software aumente en los próximos años debido al crecimiento de la demanda y las expectativas de los jugadores por los mundos virtuales [9].
2. **Optimización del Desarrollo:** Ejecutivos de la industria de videojuegos creen que en un plazo de 5 a 10 años, la IA podría gestionar más de la mitad del desarrollo de juegos, lo que sugiere una adopción considerable de la IA en la creación de mundos virtuales y mejora de la calidad de los juegos [11].
3. **Atención de los Desarrolladores:** Las herramientas de generación de imágenes mediante IA están captando la atención de los desarrolladores a nivel global, lo que indica

un interés creciente en la generación de mundos virtuales mediante IA [12].

4. **Generación Rápida de Activos 3D:** Las tecnologías de IA generativa permiten la creación rápida de activos 3D para mundos virtuales, lo que puede ser crucial para escalar la producción de entornos de juego y realidad virtual [14].

1.1.2. Motivación Personal

Mi interés por los videojuegos y la Inteligencia Artificial generativa se fusiona en la motivación detrás de esta propuesta. Los videojuegos, con sus mundos virtuales inmersivos, siempre han estimulado mi imaginación. La Inteligencia Artificial generativa, por otro lado, me ha proporcionado una perspectiva técnica para hacer realidad estas ideas creativas. Esta investigación representa un emocionante cruce entre estos dos ámbitos. Veo la oportunidad de contribuir al desarrollo de entornos virtuales más realistas, dinámicos y adaptables, a través de la aplicación de código y modelos entrenados.

1.1.3. Objetivos

1.1.3.1. Objetivos principales

1. Desarrollar un modelo de IA para crear mapas de altura y mapas de colores (datos ecológicos) a partir de descripciones textuales.
2. Contribuir con una herramienta que simplifique y personalice la creación de entornos virtuales, mejorando su accesibilidad y diversidad.

1.1.3.2. Objetivos secundarios

1. Emplear los mapas generados para construir un modelo tridimensional con métodos procedimentales.
2. Evaluar la eficacia de la combinación de datos (mapas de altura y colores) en la generación de ecosistemas realistas.

1.1.3.3. Alcance

Este trabajo se enfocará en la generación de terrenos y ecosistemas a partir de descripciones textuales y datos específicos. Se llevará a cabo una evaluación cualitativa y cuantitativa de los resultados, pero no se incluirá la implementación práctica en aplicaciones de videojuegos o simulaciones en esta etapa.

1.1.4. Metodología

1.1.4.1. Estrategia de investigación

La estrategia de investigación se basará en un enfoque mixto, combinando la investigación cuantitativa y cualitativa para proporcionar una comprensión profunda y multidimensional del problema. Utilizaremos análisis cuantitativos para evaluar la eficacia, precisión y realismo de los datos generados. Además, la investigación cualitativa se empleará para entender la aplicabilidad y utilidad de los datos generados en contextos prácticos.

1.1.4.2. Metodología de trabajo

Se ha optado por desarrollar un nuevo producto, específicamente, un modelo de IA que puede generar mapas de altura y de colores basados en descripciones textuales y datos ecológicos y geográficos, abordando directamente la problemática identificada: la necesidad de una herramienta eficiente y personalizable para crear terrenos realistas.

Implementaremos la metodología Kanban, una forma de programación ágil que se centra en la entrega continua y la eficiencia del proceso. Kanban facilitará la flexibilidad para adaptarse a los descubrimientos y tareas emergentes durante el proceso de investigación y desarrollo.

1.1.5. Planificación

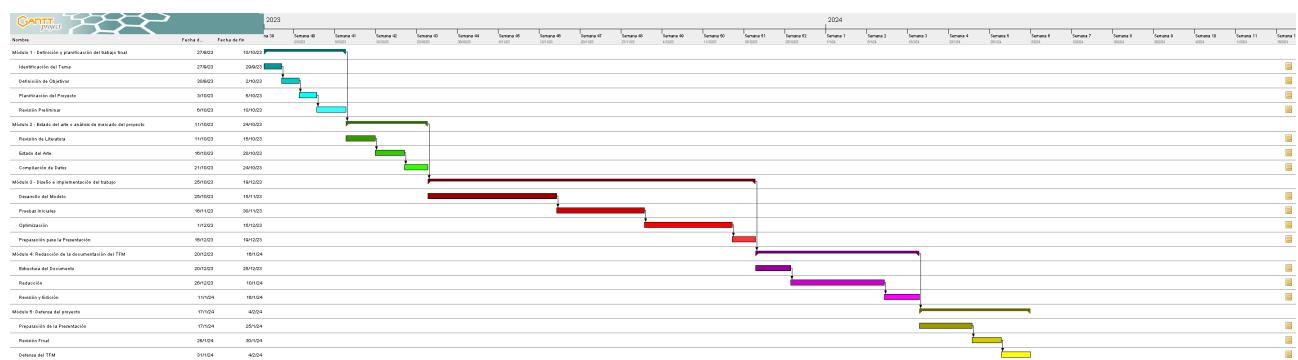


Figura 1.1: Planificación Gantt

Capítulo 2

Estado del Arte

2.1. Introducción

En el ámbito de la industria del videojuego (y otros sectores), impulsada por un crecimiento exponencial en los últimos años, la generación de contenido virtual de alta calidad se ha convertido en un factor crucial para la satisfacción de los jugadores y el éxito comercial. Uno de los desafíos clave en esta industria es la creación de entornos tridimensionales y ecosistemas realistas que no solo sean visualmente impresionantes, sino también relevantes para la narrativa y la experiencia de juego. En este contexto, la inteligencia artificial (IA) ha emergido como una herramienta fundamental para abordar este desafío, al permitir la generación automatizada de contenido virtual.

En este apartado, se presenta un análisis exhaustivo del estado del arte en relación con la generación de terrenos tridimensionales y ecosistemas realistas asistida por IA, con un enfoque particular en cómo el aprendizaje profundo y la interpretación de descripciones textuales están transformando la forma en que se aborda este problema. Se examinan los avances actuales y las metodologías más recientes, destacando la relevancia y la aplicación de la IA en la creación de datos topográficos y biomas precisos y visualmente cautivadores. A lo largo de esta revisión, se presentarán investigaciones y desarrollos destacados que han contribuido significativamente a este campo en constante evolución.

1. **Conditional GANs:** El artículo titulado "Generating Terrain Data for Geomorphological Analysis by Integrating Topographical Features and Conditional Generative Adversarial Networks." explora la viabilidad de generar datos topográficos realistas para el análisis geomorfológico a través de un modelo basado en aprendizaje profundo, centrándose particularmente en datos de formas de terreno "loess". El modelo, llamado Terrain-CGAN, utiliza Redes Generativas Adversariales Condicionales (CGANs) para simular datos to-

pográficos basados en diversas características topográficas. El rendimiento del modelo se evalúa comparando los datos generados con datos de referencia, enfocándose en similitudes en las distribuciones de elevación y pendiente. Los resultados sugieren que los datos topográficos generados se asemejan estrechamente a los datos de referencia en términos de elevación y pendiente, aunque se reconocen algunas inexactitudes en los detalles del terreno. Los resultados resaltan la utilidad potencial de los datos simulados para apoyar la investigación geográfica y de terrenos, especialmente cuando se requieren conjuntos de datos grandes. Además, el estudio subraya la importancia de seleccionar cuidadosamente las características de entrada para mejorar la calidad y precisión de los datos topográficos generados. El modelo Terrain-CGAN muestra promesa en la generación de datos topográficos para formas de terreno "loess", y los autores sugieren que el marco podría adaptarse para otras áreas de formas de terreno. [4].

2. **AutoBiomes: Generación Procedimental de Paisajes Multi-Bioma:** En el ámbito de la generación procedural de terrenos, este artículo introduce el sistema AutoBiomes, apto para generar vastos terrenos de variados y realistas biomas. Fusionando metodologías sintéticas, basadas en física y en ejemplos, AutoBiomes combina estas con modelos de elevación digital (DEMs) y una simulación climática simplificada. Enriquecido por una característica única de colocación de activos, el sistema asegura detalladas distribuciones de múltiples objetos a lo largo del paisaje. Aunque AutoBiomes no parece incorporar tecnologías de aprendizaje automático, su relevancia para este proyecto centrado en IA es evidente. Ambos trabajos buscan generar mapas de altura creíbles, subrayando el potencial interdisciplinario entre la generación procedural y las metodologías impulsadas por redes neuronales. [1].
3. **Aprendizaje Profundo aplicado a la Generación de Contenido Procedimental:** Este artículo explora el uso del aprendizaje profundo en la generación de contenido para videojuegos. Se analiza la historia de la generación de contenido procedural en los videojuegos y la integración reciente de técnicas de aprendizaje profundo. El artículo aborda diversos métodos de aprendizaje profundo, como GANs, VAEs y LSTMs, utilizados para generar contenido de juegos. Se enfatiza la necesidad de adaptar cuidadosamente estos métodos para que se ajusten a los requisitos específicos de los juegos. El artículo también destaca las aplicaciones más amplias de la generación de contenido más allá de los videojuegos. En conclusión, se anticipa un rápido progreso en este campo y se sugieren direcciones futuras de investigación, como la generación de contenido de juegos con profundidad de habilidad y la adaptación en tiempo real basada en las preferencias de los jugadores. [2].

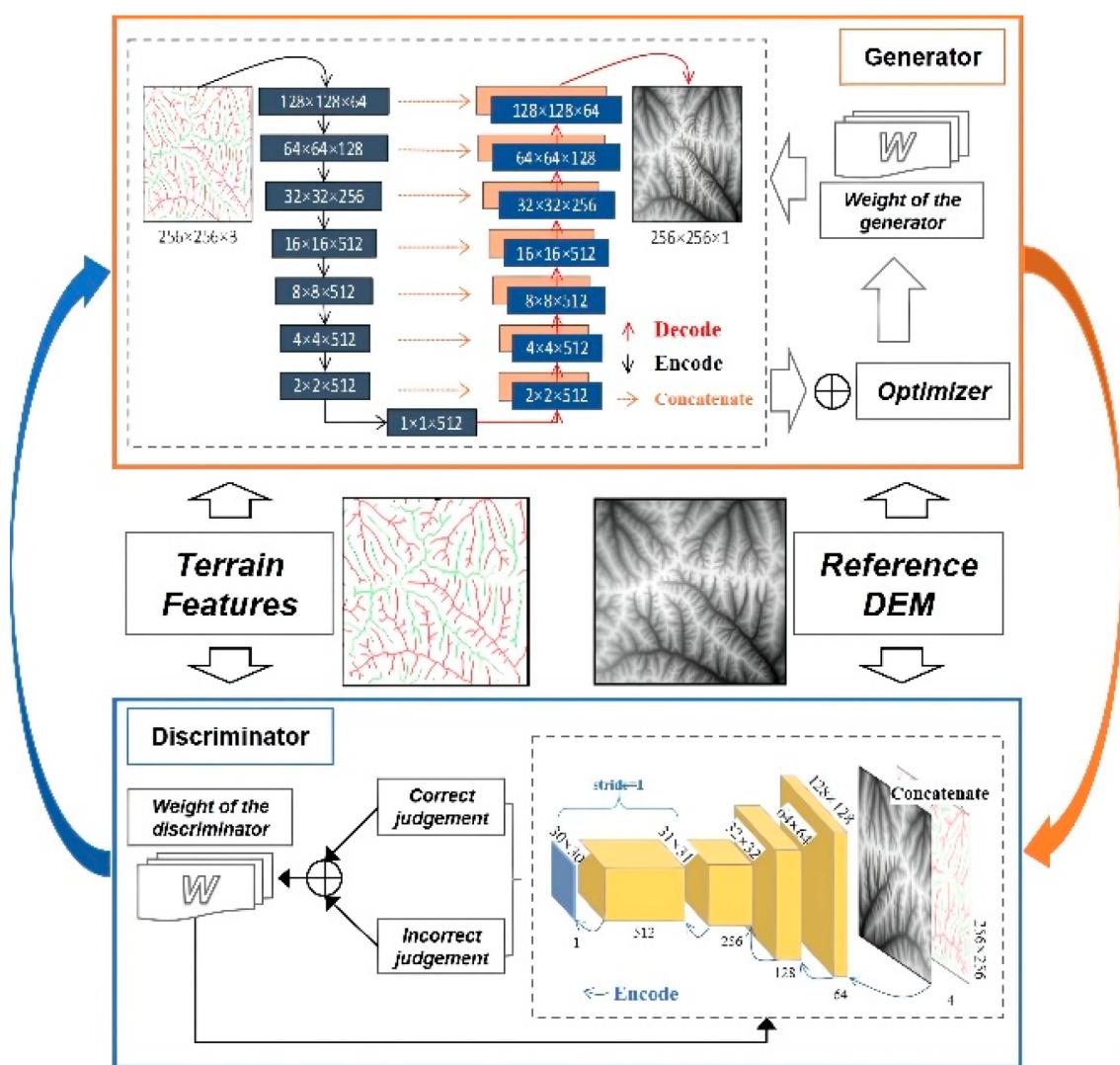


Figura 2.1: Estructura de Terrain-CGAN.

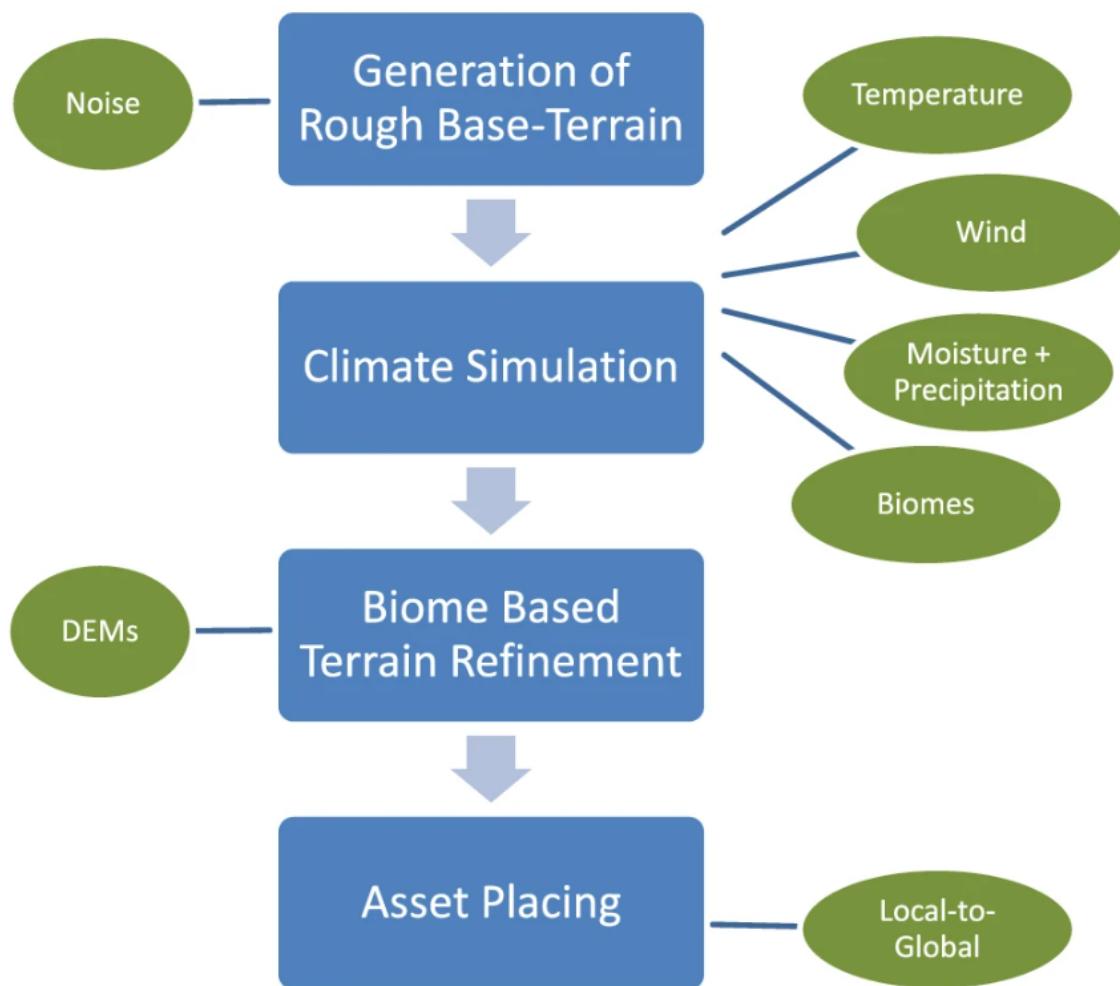


Figura 2.2: Proceso de generación de terrenos.

4. 3D-GPT: Generación de Mundos 3D a partir de Descripciones Textuales:

El artículo presenta 3D-GPT, un marco de trabajo que aprovecha modelos de lenguaje grandes (LLMs, por sus siglas en inglés) para simplificar y mejorar el proceso de modelado 3D en la era del meta-verso. Divide la tarea de modelado 3D en componentes más pequeños y utiliza LLMs como agentes resolutorios de problemas, mejorando las descripciones de las escenas y empleando la generación procedimental basada en instrucciones de texto. Este enfoque reduce el esfuerzo manual, mejora la colaboración con los usuarios e integra de manera fluida con software 3D como Blender. El artículo muestra el potencial de los LLMs en la generación de contenido 3D a través de experimentos empíricos, lo que lo convierte en una referencia relevante para la investigación en el metaverso y el modelado 3D. [5].

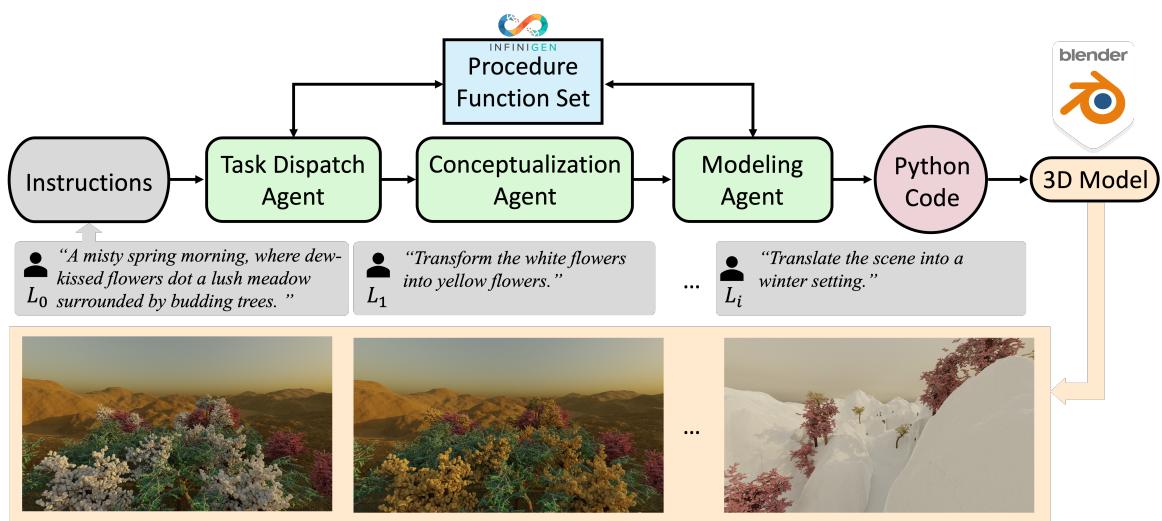


Figura 2.3: 3D-GPT emplea LLM como un sistema multiagente con tres agentes colaborativos para la generación de procedimientos 3D.

Capítulo 3

Fundamentos Teóricos

3.1. Redes Generativas Antagónicas (GANs)

Las Redes Generativas Antagónicas (GANs, por sus siglas en inglés) son una clase de marcos de aprendizaje automático introducidos por Ian Goodfellow en 2014. Se componen de dos modelos distintos: un generador y un discriminador. El generador crea imágenes 'falsas' que se parecen a las imágenes de entrenamiento, mientras que el discriminador evalúa imágenes y determina si son reales (provenientes de los datos de entrenamiento) o falsas (creadas por el generador). Durante el entrenamiento, el generador intenta mejorar generando falsificaciones cada vez más convincentes, y el discriminador trata de volverse mejor en detectar estas falsificaciones. El equilibrio se alcanza cuando el generador produce falsificaciones perfectas, y el discriminador tiene una probabilidad del 50 % de acertar si una imagen es real o falsa [6].

3.1.1. Arquitectura y Funcionamiento

La arquitectura de las GANs se basa en la interacción entre dos componentes clave: el generador y el discriminador.

- **El Generador (G):** Este componente utiliza un vector de espacio latente, z , tomado de una distribución normal estándar, para generar datos. La función del generador, $G(z)$, mapea el vector latente al espacio de datos con el objetivo de estimar la distribución de los datos de entrenamiento (p_{data}) y generar muestras falsas de esa distribución estimada (p_g).
- **El Discriminador (D):** Por su parte, el discriminador, $D(x)$, es una red que emite la probabilidad de que una entrada x (una imagen) provenga de los datos de entrenamiento en lugar de ser generada por G . Para imágenes, esta entrada es de tamaño 3x64x64. D se

entrena para asignar una alta probabilidad a los datos reales y baja a los generados por G .

En el proceso de entrenamiento, D y G juegan un juego de minimax, donde D intenta maximizar la probabilidad de clasificar correctamente los datos reales y falsos ($\log D(x)$), y G intenta minimizar la probabilidad de que D prediga que sus salidas son falsas ($\log(1 - D(G(z)))$). La función de pérdida de GAN se basa en esta dinámica. [6, 8].

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

Un ejemplo específico de GAN es la DCGAN (Red Generativa Antagónica Convolucional Profunda), que utiliza capas convolucionales en el discriminador y capas convolucionales-transpuestas en el generador. En las DCGANs, el discriminador consiste en capas convolucionales con normalización por lotes y activaciones LeakyReLU, mientras que el generador usa capas convolucionales-transpuestas, normalización por lotes y activaciones ReLU [6, 8].

3.1.2. Red Generativa Antagónica Convolucional Profunda (DCGANs)

Las DCGANs, como extensión de las GANs tradicionales, presentan diferencias clave que las hacen más aptas para trabajar con imágenes:

- **Convoluciones Estratificadas y Transpuestas:** A diferencia de las GANs tradicionales, las DCGANs reemplazan las capas de agrupamiento con convoluciones estratificadas en el discriminador y convoluciones transpuestas en el generador. Esta modificación facilita la generación de imágenes más detalladas y de mayor resolución.
- **Normalización por Lotes:** Tanto el generador como el discriminador en las DCGANs incorporan normalización por lotes. Esta técnica ayuda a estabilizar el aprendizaje y previene problemas como el desvanecimiento o explosión de gradientes, comunes en las GANs tradicionales.
- **Eliminación de Capas Ocultas Totalmente Conectadas:** En las DCGANs, se omiten las capas ocultas totalmente conectadas para estructuras más profundas, permitiendo un manejo más eficiente de diferentes tamaños de entrada y salida, lo cual es crucial al trabajar con imágenes.
- **Funciones de Activación Específicas:** Las DCGANs utilizan ReLU en el generador (excepto en la capa de salida, donde se utiliza Tanh) y LeakyReLU en el discriminador. Estas funciones de activación mejoran el flujo de gradientes durante el entrenamiento, lo que se traduce en una generación de imágenes de mejor calidad.

Estas diferencias hacen que las DCGANs sean más efectivas y estables para la generación de imágenes, especialmente en tareas que requieren una alta fidelidad visual, como la generación de mapas de alturas o imágenes realistas [6, 3, 7].

3.2. Pix2Pix: Entendiendo las Redes Condicional GANs

Las Redes Generativas Antagónicas Condicionales (Conditional GANs, o cGANs) son una extensión de las redes GAN, que permiten controlar el tipo de imagen generada. El modelo Pix2Pix es una implementación específica de cGAN donde la generación de una imagen está condicionada a una imagen dada. A diferencia de las GANs tradicionales, que generan imágenes a partir de un vector de distribución aleatoria sin ninguna condición aplicada, las cGANs pueden ser entrenadas para realizar tareas específicas de transformación de imágenes [13, 16].

3.2.1. Arquitectura del Modelo Pix2Pix

El modelo Pix2Pix utiliza una arquitectura basada en U-Net para su generador y un discriminador tipo PatchGAN [13]. El discriminador toma una imagen del dominio fuente y otra del dominio objetivo, y predice la probabilidad de que la imagen del dominio objetivo sea una versión real o generada de la imagen fuente [10].

3.2.1.1. Función de Pérdida Compuesta

El modelo de discriminador se entrena de manera independiente, minimizando la probabilidad logarítmica negativa de identificar imágenes reales y falsas, aunque condicionada por una imagen fuente. El modelo generador se entrena utilizando tanto la pérdida adversaria del modelo discriminador como la pérdida L1 o la diferencia absoluta media de píxeles entre la traducción generada de la imagen fuente y la imagen objetivo esperada [10].

3.2.2. Aplicaciones del Pix2Pix GAN

El GAN Pix2Pix ha sido demostrado en una variedad de tareas interesantes de traducción de imagen a imagen, incluyendo la conversión de etiquetas semánticas a fotografías, de mapas a fotos aéreas, y de bocetos a fotos, entre otros [13, 15].

3.2.3. Entrenamiento y Preprocesamiento

Para entrenar el modelo Pix2Pix, se requiere un conjunto de datos de imágenes apareadas que contengan imágenes fuente y objetivo. Por ejemplo, para la tarea de texturizar mapas,

se pueden usar imágenes de satélite y mapas correspondientes [10]. Estas imágenes se cargan, dividen y escalan adecuadamente antes de alimentarlas al modelo. El entrenamiento del modelo implica el ajuste de las redes generadora y discriminadora, utilizando una función de pérdida que equilibra la generación de imágenes creíbles y la precisión de la traducción de la imagen fuente [15].

Capítulo 4

Metodología y Desarrollo

4.1. Implementación de la DCGAN para Mapas de Altura

Explicaremos la implementación realizada de la DCGAN para llevar a cabo la primera parte de este proyecto: la generación de mapas de altura (heightmap) realistas.

(WIP) https://pytorch.org/tutorials/beginner/dcgan_faces_tutorial.html#weight-initialization->Fuente

4.1.1. Preprocesado de Datos

[...]

4.1.2. Inicialización de Pesos

En el paper de DCGAN, los autores especifican que todos los pesos del modelo se inicializarán aleatoriamente a partir de una distribución normal con media = 0, stdev = 0,02. La función *weights_init* toma un modelo inicializado como entrada y reinicializa todas las capas convolucionales, de transposición convolucional y de normalización por lotes para cumplir con este criterio.

[Insertar código]

4.1.3. Generador

El generador, G , se encarga de convertir un vector del espacio latente en imágenes RGB, con el mismo tamaño que las utilizadas en el entrenamiento, es decir, $3 \times 64 \times 64$. Este proceso

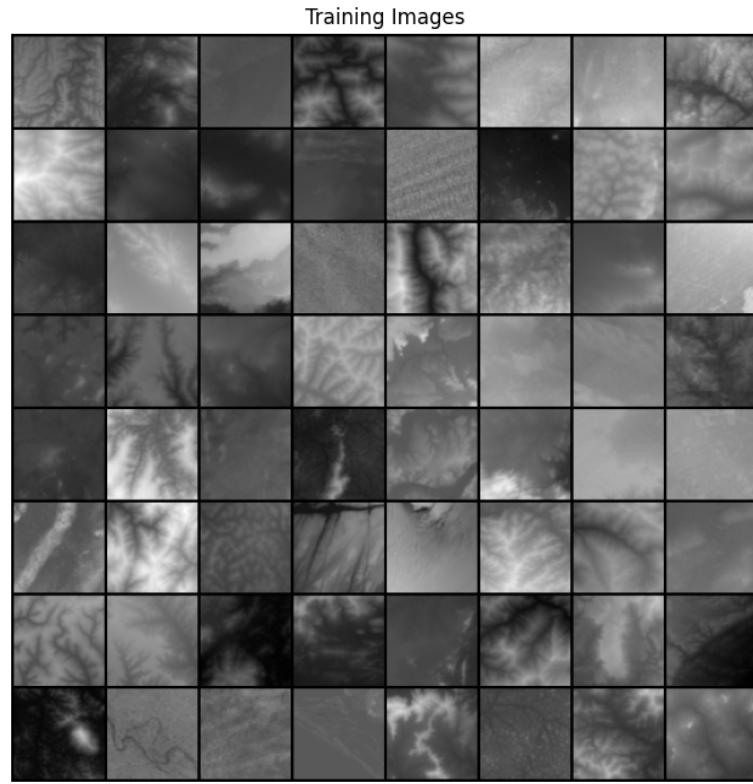


Figura 4.1: Datos de Entrenamiento

se realiza a través de capas convolucionales bidimensionales de transposición, cada una acompañada de una capa de normalización por lotes 2D y una activación ReLU. La salida de G se procesa mediante una función tanh para ajustarla al rango de los datos de entrada $[-1, 1]$. Es importante destacar el uso de las capas de normalización por lotes después de la convolución-transposición, ya que son clave para la eficacia del entrenamiento en el enfoque DCGAN. A continuación, se presenta una imagen del generador de acuerdo con el paper DCGAN.

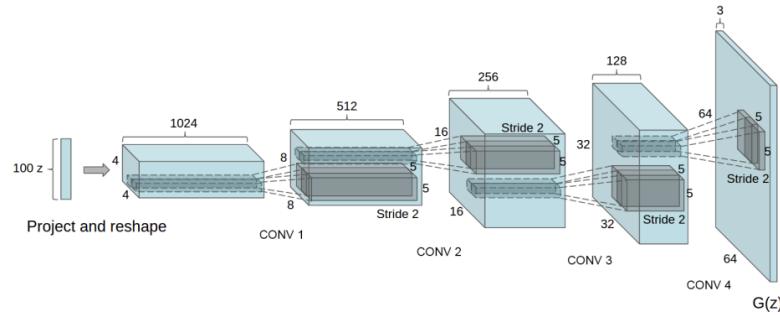


Figura 4.2: Estructura de un Generador

[Insertar código]

4.1.4. Discriminador

El discriminador, D , es una clasificación binaria red que toma una imagen como entrada y genera una probabilidad escalar que la imagen de entrada sea real (en lugar de falsa). Aquí D recibe una imagen de entrada de $3 \times 64 \times 64$, la procesa a través de una serie de Conv2d, BatchNorm2d y LeakyReLU, y genera la probabilidad final a través de una función de activación sigmoide.

[Insertar código]

4.1.5. Funciones de pérdida y optimizadores

[...]

4.1.6. Entrenamiento

[...]

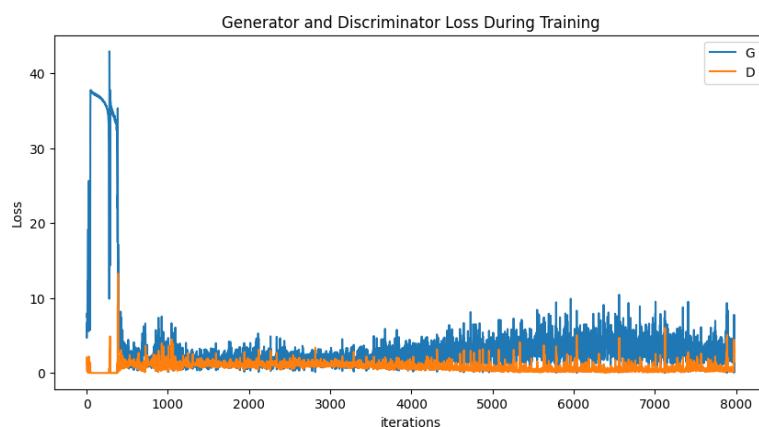


Figura 4.3: Métricas

4.1.7. Resultados y Evaluación

[...]

4.2. Implementación de Pix2Pix para Texturización

Explicaremos la implementación realizada del Pix2Pix para la segunda parte del proyecto: la generación de texturas a partir de un mapa de altura.

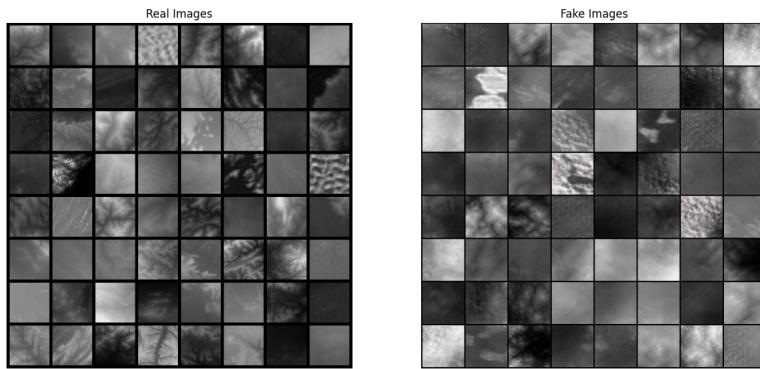


Figura 4.4: Resultados

(WIP) <https://medium.com/@Skpd/pix2pix-gan-for-generating-map-given-satellite-images-using-pytorch-6e50c318673a> -_zFuente

4.2.1. Preprocesado de Datos

(WIP) <https://www.kaggle.com/datasets/tpapp157/earth-terrain-height-and-segmentation-map-images/data> -_zDataset utilizado [Explicar el preprocesado realizado]

[...]

4.2.2. Generador

La arquitectura de un Generador, en este caso, Pix2Pix, es un codificadora-decodificadora (encoder-decoder), muy parecida a la arquitectura de una U-Net. En esta última, tanto el codificador como el decodificador son imágenes especulares entre sí.

En nuestro modelo, el codificador tiene seis bloques de convolución. Cada bloque duplica los filtros del anterior y reduce a la mitad las dimensiones de la imagen, por ejemplo, de 64x64 a 32x32. Estos bloques incluyen una capa convolucional, normalización por lotes y la función de activación LeakyReLU.

Luego, hay una capa de cuello de botella que no incrementa el número de filtros, sirviendo de conexión entre el codificador y el decodificador.

El decodificador, siendo una imagen espejo del codificador, duplica las dimensiones de cada bloque de imagen. Sin embargo, hay una diferencia en la entrada del decodificador: se concatena la salida del bloque de convolución anterior del decodificador con la salida correspondiente del codificador, como se muestra con una flecha gris en la figura.

[Insertar código]

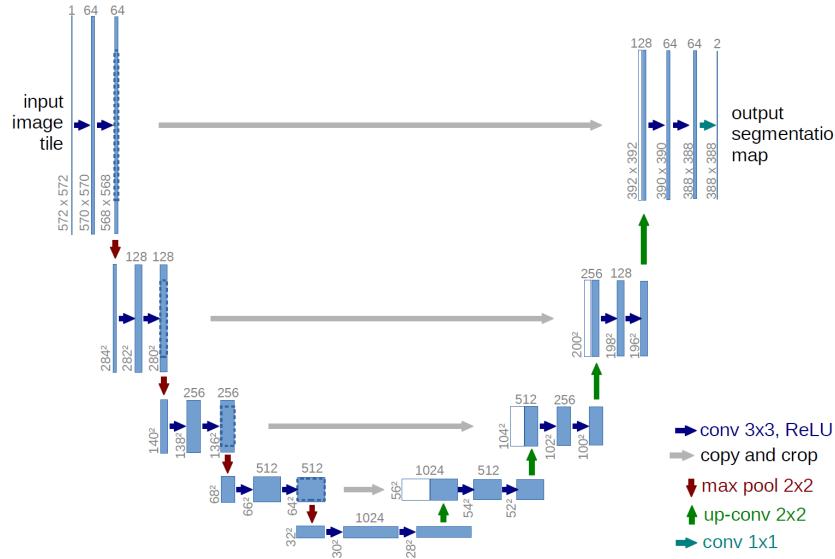


Figura 4.5: Ejemplo de una arquitectura U-Net básica (Ref:- <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/u-net-architecture.png>

4.2.3. Discriminador

El Discriminador en la arquitectura pix2pix funciona como un clasificador para determinar si una entrada es real o falsa. Al alimentar el Discriminador, se concatenan imágenes de dos dominios diferentes, como una imagen de satélite y su correspondiente mapa.

Según el artículo de investigación, se recomienda usar cuatro bloques de convolución en el Discriminador, con un número específico de filtros en cada uno: 64, 128, 256 y 512. Cada bloque consta de una capa convolucional, una capa de normalización por lotes y una capa con la función de activación LeakyReLU.

[Insertar código]

4.2.4. Entrenamiento

[...]

4.2.5. Resultados y Evaluación

[...]

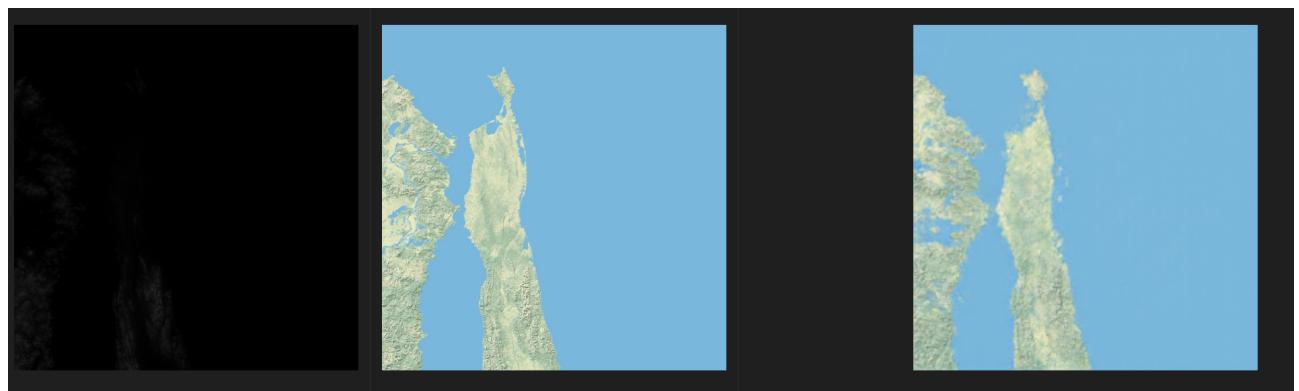


Figura 4.6: Input, Label, Output

Bibliografía

- [1] Autobiomes: procedural generation of multi-biome landscapes. 2020.
- [2] Deep learning for procedural content generation. 2020.
- [3] Implementing deep convolutional gan (dcgan) with pytorch. 2020.
- [4] Generating terrain data for geomorphological analysis by integrating topographical features and conditional generative adversarial networks. 2022.
- [5] 3d-gpt: Procedural 3d modeling with large language models. 2023.
- [6] Dcgan tutorial. 2023.
- [7] Deep convolutional gan. 2023.
- [8] Pytorch gan. 2023.
- [9] Bitkraft. How ai is helping build, and humanize, virtual worlds, 2023.
- [10] Jason Brownlee. How to develop a pix2pix gan for image-to-image translation. <https://machinelearningmastery.com/how-to-develop-a-pix2pix-gan-for-image-to-image-translation/>, 2021.
- [11] Bain & Company. How will generative ai change the video game industry?, 2023.
- [12] PC Gamer. This developer is paving the way for ai generated worlds in virtual reality, 2023.
- [13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arxiv preprint arXiv:1611.07004*, 2017.
- [14] NVIDIA. Rapidly generate 3d assets for virtual worlds with generative ai, 2023.
- [15] TensorFlow. pix2pix: Image-to-image translation with a conditional gan. <https://www.tensorflow.org/tutorials/generative/pix2pix>, 2023.

- [16] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.